Influence of Error Rate on Frustration of BCI Users

Andéol Évain, Ferran Argelaguet Sanz, Anthony Strock, Nicolas Roussel, Géry Casiez, Anatole Lécuyer

Influence of Error Rate on Frustration of BCI Users

Andéol Évain*, Ferran Argelaguet†, Anthony Strock‡,
Nicolas Roussel†, Géry Casiez¶, Anatole Lécuyer†
*Université de Rennes 1, †Inria , ‡ENS Rennes, ¶Université de Lille
*†¶{andeol.evain, fernando.argelaguet_sanz, nicolas.roussel, gery.casiez, anatole.lecuyer}@inria.fr , ‡anthony.strock@ens-rennes.fr

ABSTRACT

Brain-Computer Interfaces (BCIs) are still much less reliable than other input devices. The error rates of BCIs range from 5% up to 60%. In this paper, we assess the subjective frustration, motivation, and fatigue of BCI users, when confronted to different levels of error rate. We conducted a BCI experiment in which the error rate was artificially controlled. Our results first show that a prolonged use of BCI significantly increases the perceived fatigue, and induces a drop in motivation. We also found that user frustration increases with the error rate of the system but this increase does not seem critical for small differences of error rate. Thus, for future BCIs, we would advise to favor user comfort over accuracy when the potential gain of accuracy remains small.

Keywords

BCI; frustration; SSVEP ; fake feedback; motivation; fatigue

Categories and Subject Descriptors

H.5.2. [**Information interfaces and presentation**]: User Interfaces

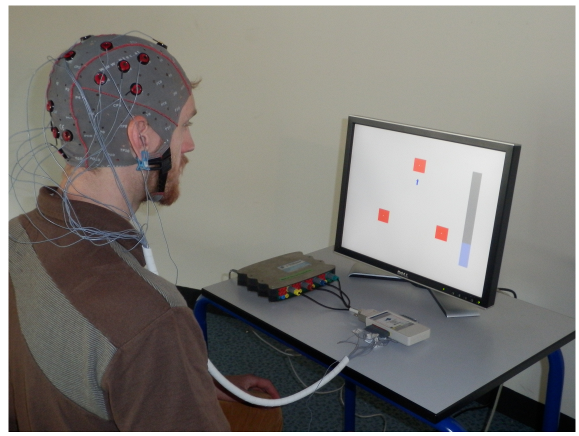

Figure 1: Experimental apparatus. A participant wearing an EEG headset tries to select one of the 3 flickering targets by focusing on it. The progression bar, visible on the right of the screen, indicates how many trials have been successful so far.

1. INTRODUCTION

Brain-Computer Interfaces (BCIs) should enable to make a dream come true: interacting with computers by the sole means of the thought. BCIs are considered as a promising approach in several application fields, such as for the assistance to disabled people [7] or the mass market of videogames [1, 9]. Since the pioneering works of the 1970s, a lot of progress has been made regarding BCI technology and the reliability of brain activity classification using electro-encephalography signals (EEG).

However, it is well-known that the error rate of BCI systems remains much higher compared to other input devices. This is probably the reason why current BCI research is mostly focused on improving the robustness of EEG signal-processing and not on studying the user experience of BCI systems. In particular, we do not know if and how the error rate can influence the user experience and the perceived fatigue. How does the user frustration correlate with the BCI system accuracy? How do frustration, motivation, or fatigue evolve with the duration of the BCI use?

In this paper, we present the results from a BCI experiment aiming at characterizing the subjective frustration, motivation and fatigue of participants while the system error rate is artificially controlled. Participants were instructed to perform a simple target selection task using a BCI (see Figure 1), and were unaware that they had no actual control over the selection task. To measure the evolution of their subjective experience, they had to fill individual questionnaires presented after each block of trials.

The remainder of this paper is organized as follows. We first present related work on BCIs and user experience. Then, we describe our experimental apparatus and methods, followed by the results of our "fake BCI" experiment. The paper ends with a general discussion and a conclusion.

2. RELATED WORK

Errors made by Brain-Computer Interfaces in the identification of brain activity are common and typically around 20% [1]. There is often a trade-off to make between the accuracy (reliability) and the speed (or time to take a decision) of a BCI system, knowing that both might impair the interaction. As stressed by [10], most BCI studies are primarily focused on lowering the error rate of BCI systems and EEG classification schemes. But several researchers have pointed out that "for general acceptance of this technology, usability and user experience will need to be taken into account when designing (BCI) systems" [1]. Several factors could indeed impair the BCI user experience such as the reliability, the delay, but also the kind of visual stimulations used [7].

For the general case of Human-Computer Interaction, it has been shown that errors in the interface are a primary cause of user frustration [3]. Other known sources of frustration are the delay [3] and hard-to-find features [3]. Frustration is often used as one dimension of the user experience. Other dimensions being, for example, the *mental demand*, *physical demand*, *performance*, or *effort* of the user [8]. Reducing user frustration is a primary concern when designing interactive systems.

In the case of interfaces based on a mouse and a keyboard, most errors are attributable to the user. For BCIs, the distinction between errors due to the system and errors due to the user is much less clear. Additionally, the range of error rate observed in BCI (typically around 20%) is very different to the one of, e.g, a mouse (less than 5%). In such conditions, it is yet unknown how big of a difference of accuracy is necessary to have a real impact on user experience. In the BCI community, it is generally assumed that the unreliability of current BCI systems could have a negative impact on user experience, and that a prolonged use of a BCI could lead to fatigue or boredom [9]. Some studies began to explore the effect of error rates for BCIs [4, 2]. However, in these studies, participants are using a keyboard as input, and are perfectly aware that errors are artificial. In order to explore the specificity of BCIs regarding user experience, we intend to reproduce the BCI context more precisely. In this paper, we report on an experiment conducted to assess specifically the influence of error rate and reliability on the perceived frustration, fatigue and motivation of BCI users.

Several techniques have been proposed to assess the user frustration and the user experience. In a recent example, a correlation was observed between some specific mouse movements and stress [11]. BCIs and brain activity can also be used as a complementary measurement, in order to qualify the *mental workload* [6] or inform about visual comfort [5]. However, the use of questionnaires remains one of the most widespread tool for evaluating the user experience [8].

3. MATERIALS AND METHODS

We aimed at assessing the influence of error rate on frustration, fatigue and motivation of BCI users. Participants were asked to perform a selection task using a BCI and wearing an EEG headset. EEG data was recorded but, unknown to the participants, it was not used during the selection task. Instead, in order to fully control the error rate of the BCI, we used a fake feedback presenting pre-determined results with an error rate varying between 5 to 50%.

3.1 Participants

Twelve participants were enrolled in this study: 7 men and 5 women, aged between 18 and 60 (mean 29, SD 12), 11 right handed and 1 left handed. All of them were naive to the purpose of the experiment and to the fact that they had no control over the selection task.

3.2 Brain-Computer Interface

The used BCI relied on the Steady-State Visually Evoked Potential (SSVEP) [12]. When the eye is stimulated at a fixed frequency (e.g. looking at a flickering GUI element) between 4 and 60 Hz, the frequency of stimulation can be observed in the activity of the visual cortex. SSVEP-based BCIs mostly use visual targets flickering at different frequencies. When the user focuses on one flickering target, the corresponding frequency is detected in EEG signals, the associated target is selected, and the corresponding command can be triggered. This method was proved very effective, with information transfer rates above 100 bpm [12]. SSVEP-based BCIs are rather robust to external noise, they require limited training, and they have relatively stable performance across users. However, the flickering stimulation can be tiring and uncomfortable for the user [7].

During the experiment, EEG data was not used, but in order to put the participants in the same condition as in a real BCI session, the EEG headset was installed normally, and the EEG data was recorded. Electrodes were positioned according to the extended 10-20 system on CPZ, POZ, OZ, IZ, O1 and O2. Additionally, a reference electrode was located on the right hear, and a ground electrode on AFZ. Signal quality was ensured using an impedance checking of each electrode.

3.3 Experimental design

Participants were sitting in front of a computer screen, wearing an EEG headset (see Figure 1). On screen, three flickering square targets were displayed. For each trial, participants were instructed to select one target (indicated with an arrow), by simply focusing on it looking away during 4 seconds. A fake feedback based on visual and auditory cues was provided at the end of each trial, indicating either a success or a failure. The visual feedback was a single word displayed at the center of screen (*success* ; *failure*). The auditory feedback was either a buzzer sound (failure) or a game-like reward sound (success). Additionally, a vertical progress bar was displayed at the right of the screen, showing the global success rate for the on-going block of trials.

Participants had to achieve blocks of 20 consecutive trials. Error rate changed depending on the block. The 4 levels of error rate were: 50% (10 errors over 20 trials), 35% (7/20), 20% (4/20), and 5% (1/20). There were 3 repetitions for each of the four conditions, for a total of 12 blocks. At the end of each block, participants had to fill a short questionnaire to gather their state. In order to avoid ordering effects, the order of the blocks was randomized. We ensured that each error rate was presented exactly once for each third of the experiment.

Each trial lasted 8 seconds (2 seconds for instructions, 4 seconds of flickering, 2 seconds of pause), and each block lasted 2 minutes and 40 seconds. The duration of the experiment was around one hour, including the time of installation and briefing.

3.4 Collected data

The questionnaires filled at the end of each block were collected for each participant. Using Likert scales, participants had to grade after each block: their frustration during the last block (*instant frustration*), their frustration since the beginning of the experiment (*global frustration*), their fatigue since the beginning of the experiment, their motivation at this stage of the experiment (*motivation*), and whether they found the interface effective during the last block (*effectiveness*). The definitions of each term (frustration, fatigue, motivation, and effectiveness) were provided on the questionnaire to ensure a good understanding of the questions. Likert scales for instant frustration, global frustration, fatigue and motivation were scaled on 7 levels : absent (1), barely perceptible (2), faintly present (3), light (4), marked (5), pronounced (6), strongly pronounced (7). Effectiveness was rated on 5 levels : strongly disagree (1), disagree (2), neither agree or disagree (3), agree (4), strongly agree (5).

4. RESULTS

The results for the ratings provided by participants at the end of the 12 blocks of trials are summarized in Table 1. The data were transformed using Aligned Rank Transform (ART) [13]. After ART, a standard two-way ANOVA analysis with Tukey post-hoc tests ($\alpha < 0.05$) was performed.

Instant frustration. The two-way ANOVA error rate and repetition vs instant frustration ratings showed a main effect on error rate ($F_{3,33}$=23.0;$p < 0.001$;η_p^2=0.67) and repetition ($F_{2,22}$=7.2;$p < 0.001$;η_p^2=0.4), no interaction effects were observed. For all levels of error rate (except between 35% and 20%) all pairwise comparisons showed significant results (all $p < 0.05$). Regarding the repetition factor, only repetitions 1 and 3 were found to be significantly different. As expected lower levels of error rate result in lower ratings of instant frustration, but also instant frustration increases with repetitions.

Global frustration. The two-way ANOVA error rate and repetition vs global frustration ratings showed a main effect on error rate ($F_{3,33}$=7.0;$p < 0.001$;η_p^2=0.39) and repetition ($F_{2,22}$=6.7;$p < 0.001$;η_p^2=0.38), no interaction effects were observed. Post-hoc tests showed that global frustration ratings at error levels of 50% and 20% are significantly higher than at 5%. Regarding the repetition factor, the global frustration for the second and third repetition was significantly higher than that of the first repetition. We could observe here that the level of global frustration raises quite fast and then seems to stabilize.

Effectiveness. The two-way ANOVA error rate and repetition vs effectiveness ratings showed only a main effect on error rate ($F_{3,33}$=17.3;$p < 0.001$;η_p^2=0.61). For all levels of error rate (except between 35% and 20%) all pairwise comparisons showed significant results (all $p < 0.05$). Thus, the participants are consistent through all the experiment and the perceived effectiveness does not seem to depend on the repetition. Then, not surprisingly, there is a strong negative correlation between error rate and perceived effectiveness of the BCI system.

Fatigue. The two-way ANOVA error rate and repetition vs fatigue ratings only showed a main effect on repetition ($F_{2,22}$=41.3;$p < 0.001$;η_p^2=0.79). Post-hoc tests showed that fatigue ratings significantly increase at each repetition.

Motivation. The two-way ANOVA error rate and repetition vs motivation ratings showed a main effect on both error rate ($F_{3,33}$=3.4;$p < 0.05$;η_p^2=0.23) and on repetition ($F_{2,22}$=6.7;$p < 0.01$;η_p^2=0.36). No interaction effects were observed. Post-hoc tests showed a significant difference between error rate levels 5% and 50%, and between the first and third repetitions. Not surprisingly, motivation strongly drops after performing a block with 50% error rate, but also lowers at the end of the experiment.

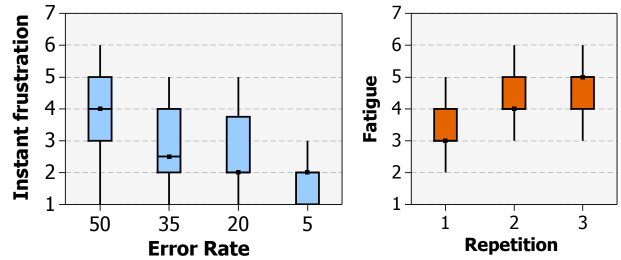

Figure 2: Experimental results. (Left) Frustration ratings as function of BCI error rate. (Right) Fatigue ratings as function of repetition.

5. DISCUSSION

Our results show that errors of the BCI system are a primary source of user frustration, as revealed by the evolution of both global and instant frustration indicators. Similar results had been observed in other HCI contexts [3], and our study confirms that BCIs are no exception. It is worth noting that our maximum error rate of 50% ends up with a median rating of *4*, i.e., a *light frustration* sensation. Thus the participants of our experiment did not express very strong feelings of frustration globally. It is possible that the user expectation regarding accuracy influences the resulting frustration. Another possibility is that if the cost of errors had been higher, the frustration feeling could have been stronger. This aspect could be specifically targeted in future work. It is also worth noting that the difference of user frustration between error rate of 20% and 35% is weak. Thus, even if accuracy seems to be an important factor of user frustration, a small improvement of this accuracy (e.g inferior to 5%) does not seem strongly influential. For the conception of BCI-based interaction systems, we would therefore advise to favor the comfort over the BCI accuracy, if the potential gain in accuracy is small, possibly even up to 15%.

Fatigue increases significantly over time, and seems to be less influenced by the error rate. The SSVEP context and the flickering targets could annoy some of the participants and generate visual fatigue after some time [12]. Besides, participants' motivation drops rapidly, and is also influenced by the error rate. This observation could be interpreted as a tendency for the participants to get discouraged quickly when confronted with a lot of errors, even if they do not feel more tired. Finally, we could observe that *effectiveness* ratings are well correlated to the error rate, suggesting that our participants were able to keep track of their performance.

At the end of the experiment, participants could answer open questions about the BCI effectiveness and their potential frustration or fatigue. Many comments confirm that the experiment is perceived as tiring, e.g., "*fatigue comes late, but all of a sudden*". Some participants complained about the duration but also the repetitiveness of the experiment, e.g., "*wish to sleep!*". Several noticed the link between failures and frustration : "*there is a bit of frustration sometimes when the system does not indicate the correct fixation, and it causes fatigue and loss of concentration*". Some of them highlighted the importance of feedback: "*the gauge gives incentive to success, the sounds (buzzer) are very frustrating*".

The two last questions posed in our post-hoc questionnaire

	Instant frustration			Global frustration			Fatigue			Motivation			Effectiveness		
Error rate	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3
50%	3	4	5	2	2	4	3	4	5	5	5	6	2	3	4
35%	2	2.5	4	2	2	4	3	4	5	5	6	6	3	4	4
20%	2	2	3.25	2	2	3.25	3	4	5	5	5	6	3	4	4
5%	1	2	2	2	2	2	3	4	5	5	6	6	3.75	4	4
Repetition	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3	Q1	M	Q3
1	1	2	4	2	2	3	3	3	4	5	6	7	3	4	4
2	2	2	4	2	2	4	4	4	5	5	5	6	3	3	4
3	2	3	4	2	2	4	4	5	5	4	5	6	2	3	4

Table 1: Participants answers to questionnaires depending on error rate and repetition. First quartile (Q1), mediane (M) and third quartile (Q3) are provided. Instant frustration, global frustration, fatigue and motivation are rated on a 7-point Likert-scale. Effectiveness is rated on a 5-point Likert-scale.

controlled if the participants had noticed the fake feedback. From their answers, it seems that five participants did not suspect anything at all, and had the impression that the responses were always consistent with their performance. In particular, the participant who performed the best considering the post-experiment EEG analysis did not have any suspicion at all. The other participants had more nuanced answers and could have been more distracted. But nobody claimed openly that he/she was sure of facing a fake feedback. Besides, it seems that they all decided to play the game and did not try to check this during the experiment : "*to really know, I would have to look into the wrong square to test if it always gives a failure, which obviously I could not do during the experiment*". Actually they were often assimilating a *failed* trial with an error of the system. For example, one participant answered "*there were times at which I fixed the correct square, but the system did not find it*". Thus, there seems to be a natural tendency to attribute success to oneself, and failure to the system.

6. CONCLUSION

We have studied the influence of error rate on the perceived experience of BCI users. To do so, we have designed a *fake* BCI experiment, in which the results of a target selection task were artificially simulated with different error rate conditions. Our results show that high error rates increase user frustration, and that this frustration accumulates over time. However, the increase in frustration does not seem critical for small differences of error rate. Thus, for future of BCIs, we would advise to favor user comfort over small improvement of accuracy. Besides the use of a fake feedback and simulated experimental conditions was barely noticed by the participants and could inspire further studies on BCI user experience in well-controlled conditions. For future work, we would like to study the influence of other potential factors such as the feedback delay, or the error cost. Then, we would like to design specific interaction techniques for BCI systems that would take into account the error rate and the cost of errors, in order to improve the BCI user experience.

7. REFERENCES

[1] D.-O. Bos, B. Reuderink, B. van de Laar, H. Gurkok, C. Muhl, M. Poel, D. Heylen, and A. Nijholt. Human-computer interaction for bci games: Usability and user experience. In *Proc. of Cyberworlds*, pages 277–281, 2010.

[2] D. P.-O. Bos, B. van de Laar, B. Reuderink, M. Poel, and A. Nijholt. Perception and manipulation of game control. In *Intelligent Technologies for Interactive Entertainment*, pages 57–66. Springer, 2014.

[3] I. Ceaparu, J. Lazar, K. Bessiere, J. Robinson, and B. Shneiderman. Determining causes and severity of end-user frustration. *International Journal of Human-Computer Interaction*, 17:333–356, 2004.

[4] P. R. Fard and M. Grosse-Wentrup. The Influence of Decoding Accuracy on Perceived Control: A Simulated BCI Study. *ArXiv e-prints*, 2014.

[5] J. Frey, A. Appriou, F. Lotte, and M. Hachet. Estimating visual comfort in stereoscopic displays using electroencephalography: A proof-of-concept. In *Proc. of INTERACT*, pages 354–362, 2015.

[6] L. M. Hirshfield, R. Gulotta, S. Hirshfield, S. Hincks, M. Russell, R. Ward, T. Williams, and R. Jacob. This is your brain on interfaces: Enhancing usability testing with functional near-infrared spectroscopy. In *Proc. of CHI*, pages 373–382, 2011.

[7] H. A. Lamti, M. M. B. Khelifa, A. M. Alimi, and P. Gorce. Effect of fatigue on ssvep during virtual wheelchair navigation. *Journal of theorical and applied information technology*, 65:1–10, 2014.

[8] L. A. Leiva, A. Sahami, A. Catala, N. Henze, and A. Schmidt. Text entry on tiny qwerty soft keyboards. In *Proc. of CHI*, pages 669–678, 2015.

[9] A. Nijholt, D. P.-O. Bos, and B. Reuderink. Turning shortcomings into challenges: Brain–computer interfaces for games. *Entertainment Computing*, 1(2):85–94, 2009.

[10] D. Plass-Oude Bos, H. Gürkök, B. van der Laar, F. Nijboer, and A. Nijholt. User experience evaluation in bci: Mind the gap! *International Journal of Bioelectromagnetism*, 13(1):48–49, 2011.

[11] D. Sun, P. Paredes, and J. Canny. Moustress: Detecting stress from mouse motion. In *Proc. of CHI*, pages 61–70, 2014.

[12] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao. Brain-computer interfaces based on visual evoked potentials. *Engineering in Medicine and Biology Magazine*, 27:64–71, 2008.

[13] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proc. of CHI*, pages 143–146, 2011.